

# Proteomic Approaches to Characterize Protein Modifications: New Tools to Study the Effects of Environmental Exposures

Daniel C. Liebler

Southwest Environmental Health Sciences Center, College of Pharmacy, University of Arizona, Tucson, Arizona, USA

Proteomics is the study of proteomes, which are the collections of proteins expressed in cells. Whereas genomes are essentially invariant in different cells in an organism, proteomes vary from cell to cell, with time and as a function of environmental stimuli and stress. The integration of new mass spectrometry (MS) methods, data analysis algorithms, and information from databases of protein and gene sequences has enabled the characterization of proteomes. Many environmental agents directly or indirectly generate reactive electrophiles that covalently modify proteins. Although considerable evidence supports a key role for protein adducts in adverse effects of chemicals, limitations in analytical technology have slowed progress in this area. New applications of liquid chromatography–tandem mass spectrometry (LC–MS–MS) now offer the potential to identify protein targets of reactive electrophiles and to map adducts at the level of amino acid sequence. Use of the data-analysis tools Sequest and SALSA (Scoring Algorithm for Spectral Analysis) together with LC–MS–MS analyses of protein digests enables the identification of modified forms of proteins in a sample. These approaches can map adducts to specific amino acids in protein targets and are being adapted to searches for protein adducts in complex proteomes. These tools will facilitate the identification of new biomarkers of chemical exposure and studies of mechanisms by which protein modifications contribute to the adverse effects of environmental exposures. **Key words:** functional genomics, mass spectrometry, protein adducts, proteomics, stress responses. *Environ Health Perspect* 110(suppl 1):3–9 (2002).

<http://ehpnet1.niehs.nih.gov/docs/2002/suppl-1/3-9liebler/abstract.html>

## One Genome, Many Proteomes

Completion of a human genome sequence will change biology and medicine in many ways (1,2). Most fundamentally, the information in a well-annotated genome sequence represents a catalog of all possible gene products in human cells. Thus, analyses of gene expression provide important information about the states of cells and about their responses to chemicals and other environmental stimuli. Genes ultimately code for proteins, which perform most of the functions of cells. The *proteome* is the protein complement of the expressed genes in a cell. Whereas the genome of an organism is essentially invariant in all its cells, proteomes vary among different cell types and also vary with time. This is because not all genes in an organism are expressed in all its cells and levels of gene expression vary with time. Moreover, any protein product of a single gene may exist in multiple forms because of posttranslational modifications, the existence of mutants, and the formation of splice variants. Indeed, it appears that many cellular proteins may exist in at least two forms at any given moment. Thus, the complexity of the proteome presents challenges not seen in gene expression analyses.

## Analytical Proteomics

Mass spectrometry (MS) is the key technology presently driving proteomics (3,4). MS analysis of peptides generated by proteolysis of protein samples provides information sufficient to identify the proteins. I present an overview of

analytical proteomics methodology in Figure 1. Protein mixtures (cell extracts, subcellular fractions, protein complexes, etc.) are first subjected to some type of separation to resolve the mixture into several fractions containing fewer components. The prototypical approach employs two-dimensional sodium dodecyl sulfate–polyacrylamide gel electrophoresis (2D SDS–PAGE), but many other protein separation methods can be used. The proteins in the separated fractions then are digested chemically or enzymatically to produce a mixture of peptides. The peptides then are analyzed by MS either with or without some prior chromatographic or electrophoretic separation. The MS data are then evaluated with the aid of database search algorithms, which correlate MS data with protein or nucleotide sequences, thus identifying the proteins represented in the mixture of peptides.

Table 1 is a summary of the features of the two principal approaches to MS analysis of peptide mixtures. Matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) MS employs laser energy to generate peptide ions from co-crystallized mixtures of peptides and ultraviolet-absorbing organic acids. MALDI–TOF analyses yield mass measurements of peptides, which can then be searched against databases with computer-assisted search algorithms to identify the proteins from which the peptides were derived. This technique, termed *peptide mass fingerprinting*, is a highly robust, useful means of rapidly identifying proteins, particularly in organisms for

which completed genome sequences or extensive protein sequence databases are available (4,5). However, this approach becomes increasingly difficult in higher organisms for which reliably annotated genome and protein sequences are unavailable. Moreover, the increasing predominance of gene paralogs (highly homologous genes) in higher organisms (6) makes it more difficult to unambiguously identify proteins by this technique (7).

Electrospray ionization–tandem mass spectrometry (ESI–MS–MS) has emerged as the second major MS technique to analytical proteomics (3). In ESI–MS–MS, peptides are ionized under high voltage at atmospheric pressure and then analyzed with either ion trap, triple quadrupole, or quadrupole–time-of-flight mass analyzers. In these analyses, the peptide ions are subjected to collision-induced dissociation, which generates fragment ions that are detected by the tandem mass analyzer. These fragmentation patterns are recorded in the resulting MS–MS spectra, which provide fingerprints that define the sequence of peptide ions. The development of the Sequest algorithm and software by Yates and Eng made possible the identification of peptides from correlation of their MS–MS spectra with virtual MS–MS data generated from protein and nucleotide sequence databases (8). Automated liquid chromatography–tandem mass spectrometry (LC–MS–MS; “data-dependent”) scanning acquires MS–MS spectra for hundreds to thousands of peptides in a single LC analysis. The combination of data-dependent scanning and Sequest automates the identification of proteins from LC–MS–MS analysis of complex peptide mixtures (9). Combination of MS–MS with reverse-phase or tandem LC (e.g., ion exchange–reverse phase) makes LC–MS–MS the most powerful approach to protein identification and characterization (10). Moreover, fragmentation of peptides in MS–MS provides unambiguous confirmation not only of sequence but also of the location

Address correspondence to D.C. Liebler, College of Pharmacy, 1703 E. Mabel St., Tucson, AZ 85721 USA. Telephone: (520) 626-4488. Fax: (520) 626-6944. E-mail: [liebler@pharmacy.arizona.edu](mailto:liebler@pharmacy.arizona.edu)

I gratefully acknowledge the contributions of H. Badghisi, J. Burr, B. Hansen, J. Jones, D. Mason, and L. Tiscareno to the tools and approaches described in this review. This work was supported in part by NIH grants ES10056 and ES06694.

Received 15 August 2001; accepted 13 November 2001.

and nature of posttranslational modifications, xenobiotic adducts, and sequence variants.

## Protein Modification in Chemical Toxicity and Carcinogenesis

In the 1940s, the Millers first discovered that chemical carcinogens underwent biotransformation to metabolites that covalently bound tissue macromolecules (11). The subsequent discoveries that binding to DNA led to molecular lesions, mutations, and cancer established a paradigm that has dominated the study of chemical carcinogenesis ever since. Nevertheless, the bulk of reactive intermediates formed in cells bind to cellular proteins. The significance of protein binding went largely unappreciated until the work of Brodie, Gillette, Jollow, Mitchell, and colleagues in the early 1970s established that protein covalent binding of acetaminophen and bromobenzene was closely associated with toxicity and suggested a causative role for protein adduction in cell injury (12–15). In the years since, a large body of data has emerged to support this view, yet the relationship between protein covalent binding and toxicity certainly is far from simple. In some cases, nontoxic analogs of toxic chemicals also covalently bind to proteins (16–18); in others, toxicity can be modulated without affecting covalent binding (19). From these data has emerged the view that adduction of some protein targets is critical to injury, whereas adduction of others is not (20–22).

Several nucleophilic residues in proteins are potential targets for reactive electrophiles. In general, the most reactive protein nucleophiles are cysteine thiols, lysine  $\epsilon$ -amines, histidine imidazoles, and protein N-terminal amines. Somewhat less nucleophilic sites that nonetheless may be targets for highly reactive electrophiles include methionine sulfur, tyrosine phenols, arginine guanidinium, serine and threonine hydroxyls, and possibly even aspartate and glutamate carboxyls (23).

Attempts to identify protein targets of reactive chemicals have been complicated by

the same difficulties that have plagued the field of protein analysis. Protein structures are complex, and proteins vary greatly in their molecular sizes and physical properties. Specific detection of adducted species is difficult because unmodified proteins are always present in excess and because of the multiplicity of protein targets for most electrophiles. Lance Pohl and his colleagues first pioneered the use of antibodies in the early 1980s to detect adducts in hepatic proteins from humans and animals treated with the volatile anesthetic halothane (24). Approximately a half-dozen hepatic proteins adducted by the halothane metabolite trifluoroacetyl chloride have been detected and subsequently identified by conventional purification methods and Edman sequencing (20). The identified targets are abundantly expressed proteins of the endoplasmic reticulum, close to the site of formation of the extremely reactive acyl halide metabolite.

Other work by the Hinson and Cohen groups over the past decade has employed antibody-based approaches to identify hepatic protein adducts of acetaminophen. These groups have collectively identified seven protein targets adducted by the *N*-acetyl-*p*-benzoquinoneimine metabolite of acetaminophen (20). These include cytosolic, mitochondrial, and microsomal proteins. Similar antibody-based detection approaches have been employed to detect protein adducts of benzoquinone (from bromobenzene) (25), diclofenac (26), *S*-(1,1,2,2-tetrafluoroethyl)-L-cysteine (27), and dichloroacetyl chloride (from trichloroethylene) (22). Other recent work by Myers et al. employed densitometric analyses of proteins on 2D SDS-PAGE whose apparent levels were altered by either *in vivo* treatment with acetaminophen or its nonhepatotoxic regioisomer 3-hydroxyacetanilide (28). However, it was not possible to determine the identities of most of the proteins affected.

Each published protein adduct identification has resulted from months or years of labor-intensive work to prepare antibodies, identify adducts, purify adducted proteins, and then characterize the proteins—all making this some of the most labor-intensive work in mechanistic toxicology. In none of these investigations has the sequence context of specific adducts been unambiguously established. These approaches also are limited by the availability, quality, and specificity of antibodies to adducts. Indeed, the

ability of the antibodies to uniformly recognize adducts on different proteins and in different sequence contexts is open to question. Progress in this field has been greatly hindered by the obstacles to identifying specific protein targets of chemicals and in mapping sequence specificity of adduction.

## Application of Mass Spectrometry to Analysis of Adducts on Known Proteins

Several studies have applied MS to study the modification of peptides and small proteins by reactive electrophiles. Early work employed fast atom bombardment–tandem mass spectrometry (FAB–MS–MS) to map the sequence locations of adducts formed by *S*-(2-chloroacetyl)glutathione, *S*-(2-chloroethyl)glutathione, and *S*-(*N*-methylcarbamoyl)glutathione in short model peptides *in vitro* (29–31). Further work from the Burlingame laboratory and subsequently from several other groups employed FAB–MS–MS or ESI–MS–MS to map xenobiotic modifications on several model proteins, including bovine serum albumin (BSA), human serum albumin, human hemoglobin, human apolipoprotein B-100, and *Escherichia coli* thioredoxin (32–38). These investigators demonstrated the feasibility of using MS analysis to pinpoint covalent modifications of proteins on specific amino acid residues within a known protein sequence. Although the ability to map modifications generated *in vitro* on known proteins represents an important advance, electrophiles in complex living systems probably modify many proteins whose identities are unknown. This requires the extension of MS-based adduct analyses to proteomes.

## Proteomic Analyses of Adducts

Proteomics approaches offer the opportunity to search for adducts in proteomes containing many proteins in both modified and unmodified forms. As noted above, a fundamental problem in such work is that the identities of the protein targets are usually unknown. An innovative approach to this problem was published by Qiu et al., who used MALDI–TOF MS to identify protein targets of the hepatotoxic analgesic drug acetaminophen (39). Total hepatic proteins from mice treated with [ $^{14}\text{C}$ ]acetaminophen were separated by 2D SDS–PAGE, and spots containing radiolabel

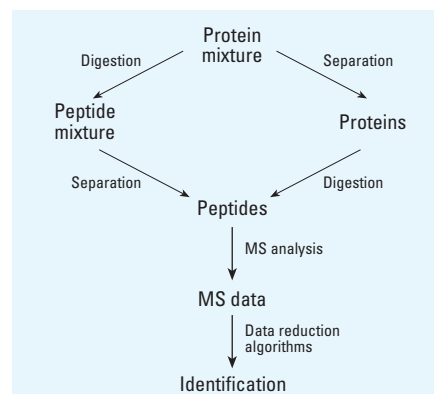


Figure 1. General scheme for proteomic analysis.

Table 1. Comparison of MALDI–TOF MS and ESI–MS–MS approaches to protein identification.

MALDI–TOF MS	ESI–MS–MS
Samples on a slide (crystalline matrix)	Samples in solution (high-performance liquid chromatography)
Spectra indicate masses of peptide ions	MS–MS spectra reveal fragmentation patterns
Protein identification by peptide mass fingerprinting	Protein identification by cross-correlation algorithms or <i>de novo</i> interpretation

were subjected to in-gel tryptic digestion and MS analysis (Figure 2). Postsource decay of peptide adduct molecular ions was used to determine sequence information for peptides derived from the proteins. The peptide masses and sequence information were used, in conjunction with protein database searching, to identify 25 protein targets, including the six that had been identified previously as targets by immunochemical methods over the past decade. This study was the first to implement a true proteomics approach to identify protein targets of reactive intermediates. Nevertheless, it is important to point out that these analyses did not, strictly speaking, identify adducts. Instead, they identified proteins that were present in two-dimensional gel spots that contained acetaminophen radiolabel. Although it is likely that the spots contained a mixture of adducted and unadducted proteins, adducts were not demonstrated on the proteins identified by sequencing. Another problem with this general approach is that radiolabel is required to select proteins from the two-dimensional gels for MS analysis. Aside from the expense and frequent synthetic difficulties of using radioisotopes, limitations in the specific activity of most useful radiochemicals will tend to bias the observer toward the detection of adducts to high-abundance proteins.

## The SALSA Algorithm

As noted above, the only MS approach that unambiguously allows the assignment of sequence location of modifications in a protein is LC-MS-MS. For LC-MS-MS analysis, the proteins in a sample must first be digested to peptides. Because one does not know which peptides in a mixture bear the modifications, the only logical approach is to collect MS-MS spectra for all the peptides in a sample, both modified and unmodified, and then determine which spectra correspond to modified peptides.

The Sequest algorithm provides a useful means of identifying proteins represented in a peptide digest (see above). However, not all the MS-MS data obtained in LC-MS-MS analyses can be evaluated successfully with

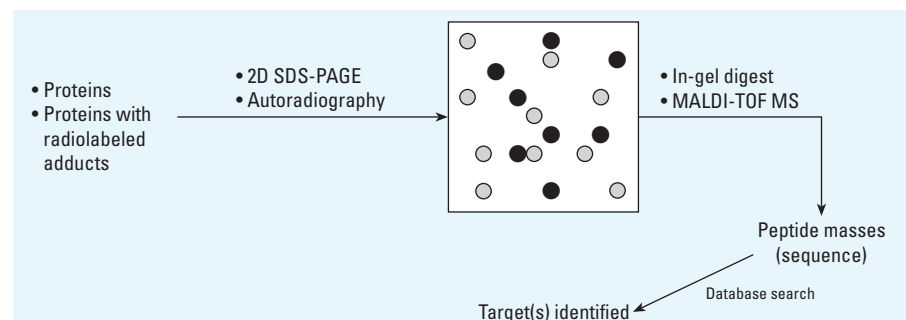
Sequest. Although Sequest can take into account some known peptide modifications on specific amino acids when matching spectra to database sequences, large or unexpected modifications or sequence variations not found in databases will yield inaccurate protein identifications. Thus, a data evaluation tool suited to detecting adduct-specific features in MS-MS spectra is needed to selectively identify spectra of modified peptides.

We have developed a novel algorithm and software called SALSA (Scoring Algorithm for Spectral Analysis) (40). The SALSA algorithm evaluates MS-MS spectra for specific, user-defined features, including product ions at specific  $m/z$  values, neutral or charged losses from singly or doubly charged precursors, and ion pairs or series (Figure 3). The MS-MS scans containing the specified features are then scored on the basis of the intensities of the detected ion signals. The scores are independent of the concentrations of the analytes in the mixture. Peptide adducts formed with the electrophiles 1,4-benzoquinone (41), 2-chloroacetyl chloride and *S*-(2-chloroacetyl)glutathione (42), 7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene (43), and dehydromonocrotaline pyrrole (40) all have been shown to display characteristic combinations of product ions, neutral or charged losses, or ion pairs that would enable detection of their MS-MS spectra with SALSA. Table 2 is a summary of characteristic product ions (PI), neutral losses (NL), charged losses (CL), and ion pairs (IP) that appear in MS-MS spectra of peptides bearing endogenous and xenobiotic-derived modifications. This list is not necessarily a complete compilation of such MS-MS features, yet it illustrates the variety of fragmentation characteristics that distinguish posttranslational modifications and xenobiotic adducts. SALSA is particularly applicable to identifying spectra of modified peptides that display multiple specific product ions, neutral losses, charged losses and ion pairs in different combinations (40).

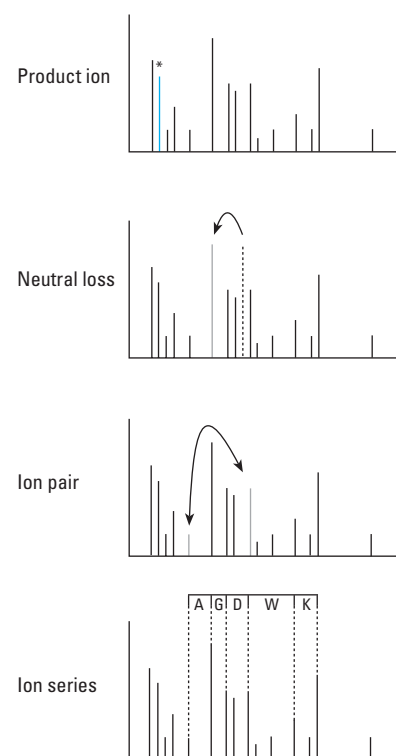
The SALSA algorithm also can search for MS-MS spectra that display a series of ions spaced by designated  $m/z$  values. These series

can correspond to b- or y-ion series that are indicative of specific peptide sequences (44). This feature is particularly valuable for finding MS-MS spectra from specific peptide sequences and can distinguish highly similar peptides of the same  $m/z$  value that have subtle sequence differences. SALSA is thus suited to the detection of MS-MS spectra displaying either specific posttranslational modifications (e.g., phosphorylation) or sequence variations (e.g., mutants or polymorphic variants).

It is important to emphasize that SALSA scores are determined by several factors, including *a*) the search strategy used, *b*) the length of the search motif, *c*) the number of ions that match the search series, and *d*) the intensities of the scored ions (44). SALSA scores do not provide an absolute measure of spectral quality or the fidelity of the match between the search motif and the MS-MS spectrum. Thus, the absolute values of SALSA scores are less important than the relative values for ranking MS-MS scans in a data set. A ranking of the MS-MS scans by SALSA score quickly identifies those MS-MS scans originating from the target peptide or its modified or variant forms. We have not attempted to statistically analyze the relationship between SALSA scores, MS-MS spectral characteristics, and peptide sequences. Depending on the search strategy used and features of individual spectra, SALSA may assign relatively high scores to spectra that do not contain the



**Figure 2.** Proteomic approach employed by Qiu et al. (32) to identify putative protein targets of [ $^{14}\text{C}$ ]acetaminophen.



**Figure 3.** MS-MS spectral features detected by the SALSA algorithm. Asterisk (\*) indicates a specific production ion. See text for discussion.



motif of interest. Thus, it is important that the user validate SALSA "hits" by inspection of the indicated spectra.

## Mapping Protein Modifications with SALSA and Sequest

Recent work in our laboratory indicates that SALSA can be used to detect modified forms of BSA, which we have employed as a model

for these studies (44). These studies analyzed commercially purchased BSA (Sigma, St. Louis, MO USA), which was not subjected to any treatment with oxidants or xenobiotics prior to analysis. Tryptic digests of BSA were analyzed by LC-MS-MS, and the data then were analyzed with Sequest and SALSA. Figure 4 depicts a summary of the data. Sequest analysis of the data identified

MS-MS spectra corresponding to 37 BSA tryptic peptides and accounting for 66.2% coverage by amino acid sequence. A SALSA analysis of the same data file was performed with ion series searches corresponding to the central sequence of each peptide. The SALSA analysis assigned significant scores to the same MS-MS spectra assigned to the peptides by Sequest. In Figure 4, the highlighted sequences indicate peptides to which MS-MS spectra were assigned by Sequest and SALSA. The first number in parentheses below each highlighted peptide corresponds to the number of MS-MS scans assigned to the sequence by Sequest; the second number corresponds to the number of MS-MS spectra assigned to the sequence by SALSA, and the third number indicates the number of modified or variant forms of that peptide sequence assigned to MS-MS spectra by SALSA. The data summary indicates that SALSA detected the MS-MS spectra for unmodified peptides also detected by Sequest. However, SALSA detected MS-MS spectra of several variant peptides not assigned by Sequest. All detected MS-MS scans assigned as variants of a target sequence displayed very strong y-ion series identity or homology (i.e., the series was displaced along the *m/z* axis) with the unmodified peptides. As indicated in Figure 4, one MS-MS scan corresponding to a variant peptide was found for eight of the peptides detected. Two variants were found for another peptide (KVPQVSTPTLVEVSR), three variants were found for two other peptides (LFTFHADICTLPDTEK and TVMENFVAFVDK), and four variants each were found for three other peptides (MPCTEDYLSLILNR, CCTESLVNR, and CCAADDKEACFAVEGPK). Inspection of the MS-MS scans corresponding to variant forms of the peptides MPCTEDYLSLILNR and CCAADDKEACFAVEGPK indicated that they were due primarily to M+16 and M+32 variants, reflecting oxidative modification at the cysteine and cysteine/methionine, respectively.

An important advantage of employing LC-MS-MS and SALSA as described in the example above is that modifications to a protein can be detected without prior knowledge of the exact chemical nature of the modifying species. In previous studies of chemical modification of proteins (see above), the investigators knew the chemical nature of the modifying species and thus could directly search the MS data for MS-MS spectra of known peptides modified by chemicals of known mass. However, unanticipated modifications would not be detectable by this approach. Moreover, some modifications by known electrophiles may yield adducts that undergo adventitious oxidation, hydrolysis, or other modifications during sample work-up.

**Table 2.** Representative MS-MS spectral characteristics of endogenous and xenobiotic-derived peptide modifications.

Modification (amino acid) <sup>a</sup>	Spectral characteristics <sup>b</sup>	Reference
Phosphorylation (S, T)	NL 98; IP 167 (pS), 181 (pT)	(62,63)
Phosphorylation (Y)	IP 243	(64)
Nitration (Y)	IP 208	(65)
Benzoquinone (C)	NL 142; IP 142, 211	(41)
Pyrrolizidine dehydropyrrole (M+135 adduct) (C)	PI 118, 120, 136; NL 135; CL 135	(40)
Pyrrolizidine dehydropyrrole (M+117 adduct) (C)	PI 118, 120; NL 117; CL 117	(40)
<i>N,N</i> -Diethylthiocarbamoyl (C)	PI 100, 72; IP 202; CL 100	(66)
Benzo[a]pyrene diol epoxide (H)	PI 303, 285, 257; NL 303; CL 303	(43)
Chrysene diol epoxide (U)	PI 279, 261,233; NL 279; CL 279	(43)
5-Methylchrysene diol epoxide (U)	PI 293, 275, 247; NL 293; CL 293	(43)
Benzo[g]chrysene diol epoxide (U)	PI 328, 311, 283; NL 328; CL 328	(43)
<i>S</i> -[2- <i>S</i> -(Cysteiny)l]acetylglutathione (C)	PI 75, 129, 178, 274, 307, 349; NL 75, 129, 273, 307, 348; IP 450	(42)
<i>S</i> -[2- <i>S</i> -(Cysteiny)l]ethylglutathione (C)	NL 75, 129; IP 436	(31,38)
<i>S</i> -Carboxymethyl (C)	IP 161; NL 60	(42)
2-Chloroacetyl (C, K, H)	IP 179 (C), 213 (H), 204 (K); NL 36, 78, 96, 114	(42)
4-Hydroxynonenal (H) <sup>c</sup>	IP 295	(37)

Abbreviations: CL, charged loss; IP, ion pair; NL, neutral loss; PI, product ion.

<sup>a</sup>Standard one-letter codes for the modified amino acids are used. "U" denotes that an unspecified amino acid was added. <sup>b</sup>Adduct-specific characteristics are those described in Figure 3. Spectral characteristics apply only to positive-ion tandem MS done by low-energy collision-induced dissociation on ion-trap and triple-quadrupole instruments.

<sup>c</sup>The adduct was reduced with sodium borohydride before analysis.

DTHKSEIAHRFKDLGEEHFGLVLIAFSQYLQCCPFDEHVKLVNELTEFAK TCVADESHAGCEK	(1,1,0)	(3,3,0)	(2,2,0)	(1,2,1)
SLHTLFGDELCK VASLR ETYGDMA DCEKQEPERNECFLSHKDDSPDLKLPDPNTLCDEFK	(1,2,1)	(1,1,0)	(1,1,0)	(1,1,0)
ADEKKFWGKLYEIA RHPYFYAPEL L YANK YNGVFOECCO AEDK GACLLPKIETMREKV LASSAR	(1,1,0)	(1,2,1)	(2,3,1)	(1,1,0)
QRLRCASIQKFGERALKAWSVARLSQKFPAEFVEVTK LVTDLT K VHKECCHGDLLECA D DRA DLAK		(2,2,0)	(1,2,1)	(1,1,0)
YICNDQDTISSK LKECCDKP LLEK SHCIAEVEKDAIPENLP L TADFAEDKDVCKNYQ EAK	(2,2,1)	(2,2,0)	(1,1,0)	
DAFLGSFLYEYSR RHPEYAVSVLLRLAKEYEATLECCAK DDPHACYSTVFDK LKHLVDEPQNLIK	(3,4,1)	(3,3,0)	(1,1,0)	(1,1,0) (3,3,0)
QNCDQFEK LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLILNR	(1,1,0)	(5,6,1)	(3,5,2)	(2,6,4)
LCVLHEK TPVSEKVT KCCTESLVNR RPCFSALTPDET VYPKAFDEKLFTFHADICTLPDTEKQIK	(1,1,0)	(1,1,0)	(1,4,4)	(1,1,0) (3,5,3)
KOTALVELLKHKPKATEEQLKTVMENFVAFVDK CCAADDKEACFAVEGPK LVVSTQ TALA	(2,2,0)	(3,6,3)	(1,7,4)	(1,1,0)

**Figure 4.** BSA tryptic peptides for which MS-MS scans were detected by Sequest and SALSA. Detected peptides are indicated by highlighted sequences. Numbers in parentheses in the format (*a,b,c*) beneath each highlighted sequence indicate *a*) the number of MS-MS scans for the indicated sequence detected by Sequest, *b*) the number of MS-MS scans for the indicated sequence detected by SALSA, and *c*) the number of MS-MS scans for variants of the indicated sequence detected by SALSA. Reproduced from Mason and Liebler (41) with permission of the American Chemical Society.

These unanticipated modifications would yield adducted peptides that would not be identified by directly searching the data for peptides of the expected mass modification. The advantage of SALSA in these situations is that it can identify MS–MS spectra for variant and modified peptides even when the exact nature of the modification is not known. This is because SALSA searching for spectra with ion series motifs will detect ion series patterns, which are at least partially conserved in spectra of modified peptide forms. Application of this approach to mapping chemical adducts on proteins is ongoing in our laboratory.

For analysis of modifications in complex proteomes, Sequest and SALSA will likely be used together to identify proteins in a sample and to map the sites of modifications. The general approach is illustrated in Figure 5. An initial analysis of the data with Sequest would successfully correlate many of the MS–MS spectra with database sequences. Even if some of the MS–MS spectra were of modified peptides and thus were not correctly identified by Sequest, other unmodified peptides from those same proteins would be identified. This initial Sequest search thus generates a list of proteins represented in the sample. Next, SALSA searches are done for the sequence motifs represented by peptides from those proteins. These sequence motif searches identify not only the MS–MS spectra of the unmodified peptides but also the MS–MS scans that have ion series homology to the unmodified peptide spectra. These spectra correspond to the modified peptides. Inspection of these spectra will then allow deduction of the mass and sequence location of each modification.

It is important to emphasize that this approach is entirely dependent on the analyst's ability to obtain MS–MS spectra of the modified forms of peptides. This is not a trivial point, because modified forms may often represent a small fraction of any particular protein. The analytical challenge then becomes obtaining MS–MS spectra of the adducted peptides in the presence of larger amounts of unmodified peptides. Adduct detection becomes even more difficult when the target protein(s) is of low abundance in mixtures containing higher abundance proteins. This is a particularly relevant issue when one considers that protein abundances may vary over about six orders of magnitude. Despite these challenges, the application of multidimensional chromatographic separations prior to or parallel with MS–MS analyses has proven highly effective for analysis of complex proteomes (10,45). Elaboration of similar approaches will help maximize the opportunity to record MS–MS spectra of lower abundance peptide adducts in complex mixtures.

## Protein Adducts as Biomarkers of Exposure to Environmental Chemicals

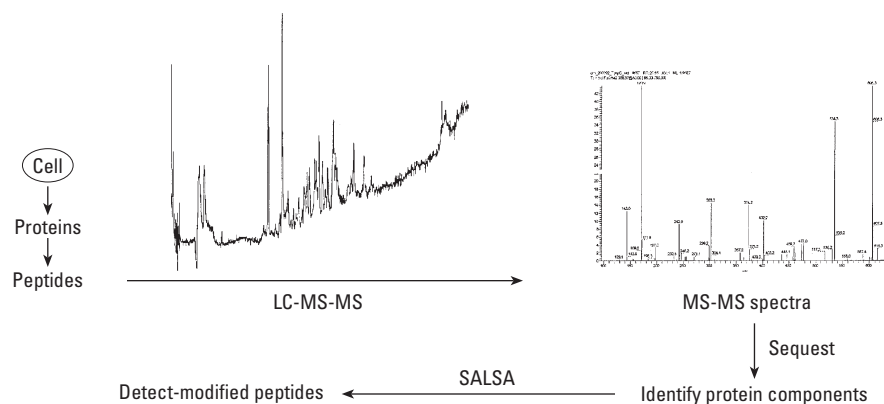
As MS and related proteomics approaches continue to evolve, more investigators will be able to identify protein adducts arising from exposure to environmental chemicals. Protein adducts may serve as markers of exposure to environmental agents. This idea is certainly not new, and a great deal of work has gone into developing methods and applications of hemoglobin and albumin adducts as markers of exposure to reactive chemicals [for reviews, see (46–48)]. Previous work on protein biomarkers has involved identifying specific, known chemical adducts on known protein targets and then developing sensitive, specific assays to detect those particular adducts.

Adaptation of proteomic approaches could lead to the discovery of new adduct biomarkers. For example, using LC–MS–MS coupled with sequence motif searching with the SALSA algorithm can identify multiple adduct variants of a target peptide based on ion series homology between the MS–MS spectra of the unadducted and different adducted peptides. We have recently demonstrated that this approach can detect multiple adducts of human hemoglobin exposed to mixtures of aliphatic epoxides *in vitro* (49). Once adducts are identified by proteomic approaches, LC–MS–MS or immunoassays may be developed to achieve quantification in samples from study populations.

## Protein Adducts as Triggers for Stress Responses

Adduction of proteins may trigger deleterious effects in living systems. This is a long-standing hypothesis that has nevertheless gone largely unexplored because of limitations in analytical technology. New techniques that identify protein targets and that map protein modifications will undoubtedly reinvigorate this field. This will permit investigators to investigate several hypotheses regarding the

mechanism(s) by which covalent modifications impact cellular functions. Certainly the dominant hypothesis over the years is that adduction results in an inhibition or loss of function of the adducted proteins (12,20,50). However, recent work suggests that oxidative or alkylation modifications to proteins trigger signaling cascades that result in activation of stress genes and phenotypic changes. Work by the Tew and Ronai groups indicates that glutathione *S*-transferase P1-1 (GSTP1-1) serves as a negative regulator of jun–N-terminal kinase 1 (JNK1) through complexation that sequesters JNK1 in an inactive form (51–54). Prooxidant stress or ultraviolet irradiation leads to dissociation of GSTP1-1 from JNK1 and to JNK1 activation (54). Similar studies indicate that the redox proteins thioredoxin (55) and glutathione *S*-transferase M1-1 (56) serve as negative regulators of apoptosis signal-regulating kinase 1 (ASK1) through a similar mechanism, in which prooxidant stress results in thioredoxin or glutathione *S*-transferase dissociation and ASK1 activation. Similarly, work by Yamamoto and colleagues revealed that the protein keap1 serves as a redox- and alkylation-sensitive switch for activation of the transcription factor NRF2, which is a principal activator of the electrophile response element. Modification of keap1 causes dissociation from NRF2, thus permitting NRF2 entry into the nucleus, complexation with other proteins, and activation of the electrophile response element (57). Studies by Stevens and colleagues have documented the ability of the prototypical alkylating agent iodoacetamide to induce the synthesis of endoplasmic reticulum (ER)-associated stress proteins, including Grp78, Grp94, and calreticulin, and the cytoplasmic stress protein Hsp70 (58–60). Other studies have provided evidence that ER proteins are prominent targets for reactive electrophiles produced from P450 enzymes, although no adducts have been mapped to specific proteins (20). The ER stress response involves activation of a complex signaling



**Figure 5.** Integrated application of LC–MS–MS, Sequest, and SALSA to identify and map protein adducts. See text for discussion.

system with “sensory” components in the ER that respond to unfolded or reduced proteins and transduce signals that result in the activation of stress gene transcriptional regulators (61). Although the mechanics of the signaling mechanisms have been elucidated in yeast and appear to be similar in higher eukaryotes, the fundamental mechanisms by which ER protein modifications trigger the cascade are obscure. This situation typifies current understanding of other stress responses. Many of the intermediate signaling events are becoming clear, but the chemical/molecular initiating events remain unexplained except in the most general terms (e.g., “protein damages” “protein oxidation,” etc.). Application of analytical proteomics approaches offers new opportunities to explicitly examine the roles of specific protein adducts as triggers for gene expression changes associated with stress.

## Conclusion

The revolution in biology spawned by completion of genome sequences for humans and other organisms offers tremendous opportunities to explore mechanisms by which environmental agents affect living systems. Proteomics approaches and gene expression technologies will fuel the emerging discipline of toxicogenomics, which describes the deleterious effects of chemicals on the expression of genes and the functions of gene products. Analytical proteomics approaches employing MS instrumentation and new data analysis tools will enable us to study the interactions of reactive chemicals with cellular proteins and proteomes. A proteomics focus is particularly relevant to studies of environmental agents, because proteins are often the initial point of contact with an organism. Moreover, effects of chemicals on proteins trigger diverse responses that contribute to injury and disease. Proteomics approaches will soon describe these mechanisms and provide a new basis for understanding cell–environment interactions.

## REFERENCES AND NOTES

- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. *Science* 291:1304–1351 (2001).
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature* 409:860–921 (2001).
- Yates JR. Mass spectrometry and the age of the proteome. *J Mass Spectrom* 33:1–19 (1998).
- Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 405:837–846 (2000).
- Fenyó D. Identifying the proteome: software tools. *Curr Opin Biotechnol* 11:391–395 (2000).
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. Comparative genomics of the eukaryotes. *Science* 287:2204–2215 (2000).
- Shevchenko A, Loboda A, Shevchenko A, Ens W, Standing KG. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal Chem* 72:2132–2141 (2000).
- Yates JR, Eng JK, McCormack AL, Schieltz D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67:1426–1436 (1995).
- Ducet A, van Ol, Eng JK, Yates JR, Aebersold R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci* 7:706–719 (1998).
- Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17:676–682 (1999).
- Miller EC, Miller JA. Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer* 47:2327–2345 (1981).
- Jollow DJ, Mitchell JR, Potter WZ, Davis DC, Gillette JR, Brodie BB. Acetaminophen-induced hepatic necrosis. II: Role of covalent binding in vivo. *J Pharmacol Exp Ther* 187:195–202 (1973).
- Jollow DJ, Mitchell JR, Zampaglione N, Gillette JR. Bromobenzene-induced liver necrosis. Protective role of glutathione and evidence for 3,4-bromobenzene oxide as the hepatotoxic metabolite. *Pharmacology* 11:151–169 (1974).
- Potter WZ, Davis DC, Mitchell JR, Jollow DJ, Gillette JR, Brodie BB. Acetaminophen-induced hepatic necrosis. 3. Cytochrome P-450-mediated covalent binding *in vitro*. *J Pharmacol Exp Ther* 187:203–210 (1973).
- Potter WZ, Thorgeirsson SS, Jollow DJ, Mitchell JR. Acetaminophen-induced hepatic necrosis. V: Correlation of hepatic necrosis, covalent binding and glutathione depletion in hamsters. *Pharmacology* 12:129–143 (1974).
- Holme JA, Hongslo JK, Bjorge C, Nelson SD. Comparative cytotoxic effects of acetaminophen (*N*-acetyl-*p*-aminophenol), a non-hepatotoxic regioisomer acetyl-*m*-aminophenol and their postulated reactive hydroquinone and quinone metabolites in monolayer cultures of mouse hepatocytes. *Biochem Pharmacol* 42:1137–1142 (1991).
- Rashed MS, Myers TG, Nelson SD. Hepatic protein arylation, glutathione depletion, and metabolite profiles of acetaminophen and a non-hepatotoxic regioisomer, 3'-hydroxyacetanilide, in the mouse. *Drug Metab Dispos* 18:765–770 (1990).
- Lau SS, Monks TJ, Gillette JR. Multiple reactive metabolites derived from bromobenzene. *Drug Metab Dispos* 12:291–296 (1984).
- Nelson SD, Pearson PG. Covalent and noncovalent interactions in acute lethal cell injury caused by chemicals. *Annu Rev Pharmacol Toxicol* 30:169–195 (1990).
- Cohen SD, Pumford NR, Khairallah EA, Boekelheide K, Pohl LR, Amouzadeh HR, Hinson JA. Selective protein covalent binding and target organ toxicity. *Toxicol Appl Pharmacol* 143:1–12 (1997).
- Hinson JA, Roberts DW. Role of covalent and noncovalent interactions in cell toxicity: effects on proteins. *Annu Rev Pharmacol Toxicol* 32:471–510 (1992).
- Pumford NR, Halmes NC, Hinson JA. Covalent binding of xenobiotics to specific proteins in the liver. *Drug Metab Rev* 29:39–57 (1997).
- Guengerich FP, Liebler DC. Enzymatic activation of chemicals to toxic metabolites. *CRC Crit Rev Toxicol* 14:259–307 (1985).
- Satoh H, Fukuda Y, Anderson DK, Ferrans VJ, Gillette JR, Pohl LR. Immunological studies on the mechanism of haloethane-induced hepatotoxicity: immunohistochemical evidence of trifluoroacetylated hepatocytes. *J Pharmacol Exp Ther* 233:857–862 (1985).
- Rombach EM, Hanzlik RP. Detection of benzoquinone adducts to rat liver protein sulfhydryl groups using specific antibodies. *Chem Res Toxicol* 10:1407–1411 (1997).
- Hargus SJ, Martin BM, George JW, Pohl LR. Covalent modification of rat liver dipeptidyl dipeptidase IV (CD 26) by the nonsteroidal anti-inflammatory drug diclofenac. *Chem Res Toxicol* 8:993–996 (1995).
- Bruschi SA, West KA, Crabb JW, Gupta RS, Stevens JL. Mitochondrial HSP60 (P1 protein) and a HSP70-like protein (mortalin) are major targets for modification during S-(1,1,2,2-tetrafluoroethyl)-L-cysteine-induced nephrotoxicity. *J Biol Chem* 268:23157–23161 (1993).
- Myers TG, Dietz EC, Anderson NL, Khairallah EA, Cohen SD, Nelson SD. A comparative study of mouse liver proteins arylated by reactive metabolites of acetaminophen and its nonhepatotoxic regioisomer, 3'-hydroxyacetanilide. *Chem Res Toxicol* 8:403–413 (1995).
- Liebler DC, Latwies DG, Reeder TC. S-(2-Chloroacetyl)-glutathione, a reactive glutathione thioester and a putative metabolite of 1,1-dichloroethylene. *Biochemistry* 27:3652–3657 (1988).
- Pearson PG, Slatter JG, Rashed MS, Han DH, Baillie TA. Carbamylation of peptides and proteins *in vitro* by S-(*N*-methylcarbamoyl)glutathione and S-(*N*-methylcarbamoyl)cysteine, two electrophilic S-linked conjugates of methyl isocyanate. *Chem Res Toxicol* 4:436–444 (1991).
- Erve JC, Deinzer ML, Reed DJ. Alkylation of oxytocin by S-(2-chloroethyl)glutathione and characterization of adducts by tandem mass spectrometry and Edman degradation. *Chem Res Toxicol* 8:414–421 (1995).
- Qiu Y, Burlingame AL, Benet LZ. Mechanisms for covalent binding of benoxaprofen glucuronide to human serum albumin. Studies by tandem mass spectrometry. *Drug Metab Dispos* 26:246–256 (1998).
- Zia-Amirhosseini P, Ding A, Burlingame AL, McDonagh AF, Benet LZ. Synthesis and mass-spectrometric characterization of human serum albumins modified by covalent binding of two non-steroidal anti-inflammatory drugs: tolmetin and zomepirac. *Biochem J* 311(pt 2):431–435 (1995).
- Ding A, Zia-Amirhosseini P, McDonagh AF, Burlingame AL, Benet LZ. Reactivity of tolmetin glucuronide with human serum albumin. Identification of binding sites and mechanisms of reaction by tandem mass spectrometry. *Drug Metab Dispos* 23:369–376 (1995).
- Ding A, Qjingwa JC, McDonagh AF, Burlingame AL, Benet LZ. Evidence for covalent binding of acyl glucuronides to serum albumin via an imine mechanism as revealed by tandem mass spectrometry. *Proc Natl Acad Sci USA* 90:3797–3801 (1993).
- Kaur S, Hollander D, Haas R, Burlingame AL. Characterization of structural xenobiotic modifications in proteins by high sensitivity tandem mass spectrometry. Human hemoglobin treated *in vitro* with styrene 7,8-oxide. *J Biol Chem* 264:16981–16984 (1989).
- Bolgar MS, Yang CY, Gaskell SJ. First direct evidence for lipid/protein conjugation in oxidized human low density lipoprotein. *J Biol Chem* 271:27999–28001 (1996).
- Erve JC, Barofsky E, Barofsky DF, Deinzer ML, Reed DJ. Alkylation of *Escherichia coli* thioredoxin by S-(2-chloroethyl)glutathione and identification of the adduct on the active site cysteine-32 by mass spectrometry. *Chem Res Toxicol* 8:934–941 (1995).
- Qiu Y, Benet LZ, Burlingame AL. Identification of the hepatic protein targets of reactive metabolites of acetaminophen *in vivo* in mice using two-dimensional gel electrophoresis and mass spectrometry. *J Biol Chem* 273:17940–17953 (1998).
- Hansen BT, Jones JA, Mason DE, Liebler DC. SALSA: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal Chem* 73:1676–1683 (2001).
- Mason DE, Liebler DC. Characterization of benzoquinone-peptide adducts by electrospray mass spectrometry. *Chem Res Toxicol* 13:976–982 (2000).
- Jones JA, Liebler DC. Tandem MS analysis of model peptide adducts from reactive metabolites of the hepatotoxin 1,1-dichloroethylene. *Chem Res Toxicol* 13:1302–1312 (2000).
- Harriman SP, Hill JA, Tannenbaum SR, Wishnok JS. Detection and identification of carcinogen-peptide adducts by nanoelectrospray tandem mass spectrometry. *J Am Soc Mass Spectrom* 9:202–207 (1998).
- Liebler DC, Hansen BT, Davey SW, Tiscareno L, Mason DE. Peptide sequence motif analysis of tandem MS data with the SALSA algorithm. *Anal Chem* [early publication: <http://pubs.acs.org>]
- Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19:242–247 (2001).
- Farmer PB. Studies using specific biomarkers for human exposure assessment to exogenous and endogenous chemical agents. *Mutat Res* 428:69–81 (1999).
- Skipper PL, Peng X, Soohoo CK, Tannenbaum SR. Protein adducts as biomarkers of human carcinogen exposure. *Drug Metab Rev* 26:111–124 (1994).
- Skipper PL, Tannenbaum SR. Protein adducts in the molecular dosimetry of chemical carcinogens. *Carcinogenesis* 11:507–518 (1990).

49. Bagdhis H, Liebler DC. Unpublished data.
50. Hinson JA, Pumford NR, Nelson SD. The role of metabolic activation in drug toxicity. *Drug Metab Rev* 26:395–412 (1994).
51. Wang T, Arifoglu P, Ronai Z, Tew KD. Glutathione *S*-transferase P1–1 (GSTP1-1) inhibits c-Jun NH<sub>2</sub>(sub2) terminal kinase (JNK1) signaling through interaction with the carboxyl terminus. *J Biol Chem* 276:20999–21003 (2001).
52. Yin Z, Ivanov VN, Habelhah H, Tew K, Ronai Z. Glutathione *S*-transferase p elicits protection against H<sub>2</sub>O<sub>2</sub>-induced cell death via coordinated regulation of stress kinases. *Cancer Res* 60:4053–4057 (2000).
53. Buschmann T, Yin Z, Bhoumik A, Ronai Z. Amino-terminal-derived JNK fragment alters expression and activity of c-Jun, ATF2, and p53 and increases H<sub>2</sub>O<sub>2</sub>-induced cell death. *J Biol Chem* 275:16590–16596 (2000).
54. Adler V, Yin Z, Fuchs SY, Benezra M, Rosario L, Tew KD, Pincus MR, Sardana M, Henderson CJ, Wolf CR, et al. Regulation of JNK signaling by GSTp. *EMBO J* 18:1321–1334 (1999).
55. Saitoh M, Nishitoh H, Fujii M, Takeda K, Tobiume K, Sawada Y, Kawabata M, Miyazono K, Ichijo H. Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J* 17:2596–2606 (1998).
56. Cho SG, Lee YH, Park HS, Ryoo K, Kang KW, Park J, Eom SJ, Kim MJ, Chang TS, Choi SY, et al. Glutathione *S*-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. *J Biol Chem* 276:12749–12755 (2001).
57. Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, Engel JD, Yamamoto M. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76–86 (1999).
58. van de WB, Wang Y, Asmellash S, Liu H, Zhan Y, Miller E, Stevens JL. Distinct endoplasmic reticulum signaling pathways regulate apoptotic and necrotic cell death following iodoacetamide treatment. *Chem Res Toxicol* 12:943–951 (1999).
59. Liu H, Bowes RC, III, van de WB, Silence C, Nagelkerke JF, Stevens JL. Endoplasmic reticulum chaperones GRP78 and calreticulin prevent oxidative stress, Ca<sup>2+</sup> disturbances, and cell death in renal epithelial cells. *J Biol Chem* 272:21751–21759 (1997).
60. Yu K, Chen Q, Liu H, Zhan Y, Stevens JL. Signalling the molecular stress response to nephrotoxic and mutagenic cysteine conjugates: differential roles for protein synthesis and calcium in the induction of c-fos and c-myc mRNA in LLC-PK1 cells. *J Cell Physiol* 161:303–311 (1994).
61. Kaufman RJ. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev* 13:1211–1233 (1999).
62. Gallis B, Corthals GL, Goodlett DR, Ueba H, Kim F, Presnell SR, Figeyes D, Harrison DG, Berk BC, Aebersold R, et al. Identification of flow-dependent endothelial nitric-oxide synthase phosphorylation sites by mass spectrometry and regulation of phosphorylation and nitric oxide production by the phosphatidylinositol 3-kinase inhibitor LY294002. *J Biol Chem* 274:30101–30108 (1999).
63. de CM, McCormack AL, Olson E, Ghomashchi F, Gelb MH, Yates JR, Leslie CC. Identification of phosphorylation sites of human 85-kDa cytosolic phospholipase A2 expressed in insect cells and present in human monocytes. *J Biol Chem* 271:6987–6997 (1996).
64. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* 71:2883–2892 (1999).
65. Jiao K, Mandapati S, Skipper PL, Tannenbaum SR, Wishnok JS. Site-selective nitration of tyrosine in human serum albumin by peroxynitrite. *Anal Biochem* 293:43–52 (2001).
66. Shen ML, Johnson KL, Mays DC, Lipsky JJ, Naylor S. Identification of the protein-drug adduct formed between aldehyde dehydrogenase and *S*-methyl-*N,N*-diethylthiocarbamoyl sulfoxide by on-line proteolytic digestion high performance liquid chromatography electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* 14:918–923 (2000).